STIC-ILL

From: Zeman, Mary
Sent: Friday, April 18, 2003 9:44 AM
To: STIC-ILL
Subject: References 09/721042

Importance: High

Please send a copy of the following references as soon as possible. Thank you,
Mary Z
Mary K. Zeman
Examiner, 1631
305-7133
CM1 12A17
mailbox: CM1 12D01

SO BioTechniques (1999), 27(6), 1218-1220, 1222, 1224
CODEN: BTNQDO; ISSN: 0736-6205

SO BioTechniques (1996), 20(6), 1090-1094, 1096-1097
CODEN: BTNQDO; ISSN: 0736-6205

SO NUCLEIC ACIDS RESEARCH, (1989 Nov 11) 17 (21) 8543-51.
Journal code: 0411011. ISSN: 0305-1048.

Volume 17 Number 21 1989 Nucleic Acids Research

A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and *in vitro* amplification of DNA

Wojciech Rychlik* and Robert E.Rhoads

Department of Biochemistry, Medical Sciences Building, University of Kentucky, Lexington, KY 40536, USA

Received July 31, 1989; Revised and Accepted October 6, 1989

ABSTRACT

A method is presented for choosing optimal oligodeoxyribonucleotides as probes for filter hybridization, primers for sequencing, or primers for DNA amplification. Three main factors that determine the quality of a probe are considered: stability of the duplex formed between the probe and target nucleic acid, specificity of the probe for the intended target sequence, and self-complementarity. DNA duplex stability calculations are based on the nearest-neighbor thermodynamic values determined by Breslauer *et al.* [*Proc. Natl. Acad. Sci. U.S.A.* (1986), 83: 3746]. Temperatures of duplex dissociation predicted by the method described here were within 0.4°C of the values obtained experimentally for ten oligonucleotides. Calculations for specificity of the probe and its self-complementarity are based on a simple dynamic algorithm.

INTRODUCTION

The quality of DNA sequencing data is a function of both the quality of the DNA and the choice of oligonucleotide used as primer. The latter is especially important when sequencing reactions are performed under non-stringent conditions favoring formation of imperfect duplexes. For example, DNA sequencing with T7 DNA polymerase is performed at room temperature, which is usually significantly below the duplex dissociation temperature (T_d) of the primer and DNA to be sequenced. Similarly, an important step in DNA amplification (polymarase chain reaction, PCR) performed with the thermostable *Thermus aquaticus* DNA polymerase (*Taq* polymerase) is choosing two oligonucleotides which are highly specific to the target DNA and not complementary, especially at their 3′ ends. For example, performing PCR with two 25-nt (nucleotide) primers which were complementary at only the two 3′-terminal residues yielded a 48-nt product (William Spencer, The Perkin-Elmer Corporation, personal communication). In addition to their use in sequencing, oligonucleotides are commonly used as hybridization probes in screening DNA libraries. They are especially useful when subsets of gene families occur and in diagnosis of genetic diseases (reviewed in 1).

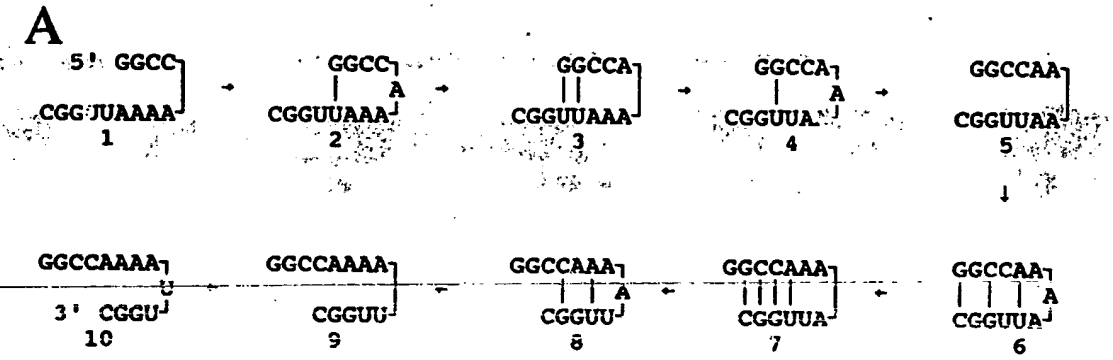
A good sequencing primer or hybridization probe should: a) form stable duplexes with the target sequence under the appropriate conditions; b) be highly specific for the intended target sequence, not base-pairing to other regions within the template; and c) not anneal to itself. The first requirement is especially important if the oligonucleotide probe is used for screening complex DNA libraries, whereas the other two are important for both screening and sequencing. A search for an oligonucleotide which would optimally meet all three of these criteria would be laborious without a computerized method.

A critical component of any such computerized method is the algorithm for determination of the duplex dissociation temperature (T_d). Some algorithms are based solely on the

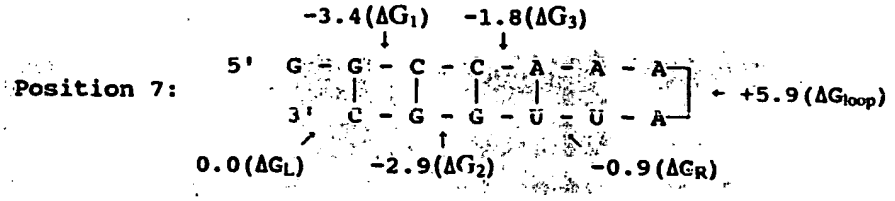
length of the oligonucleotide (reviewed in 2), but are reliable only when hybridizations are performed in tetramethylammonium salts. A more practical method of T_d determination was described by Suggs *et al.* (3):

$$T_d = 2°C \times \text{number of AT bp} + 4°C \times \text{number of GC bp.} \quad \text{(i)}$$

The most precise methods for computing helix stability, however, are based on nearest-neighbor thermodynamic parameters (4, 5). In the following report we describe a computer program, termed OLIGO, which computes T_d values based on nearest neighbor thermodynamic parameters. As a further aid to oligonucleotide selection, the program determines self-complementarity of the oligonucleotide, the presence of palindromes in the nucleic acid sequence, and the presence of alternative (non-target) sites for the oligonucleotide within the nucleic acid sequence. Experimental verification of the calculations is provided.



B



$$\Delta G = \sum_n \Delta G_n + \Delta G_L + \Delta G_R + \Delta G_i + \Delta G_{sym} + \Delta G_{loop}$$

$$\Delta G = (-3.4 - 2.9 - 1.8) - 0.0 - 0.9 + 3.4 + 0.0 + 5.9 = 0.3$$

Figure 1. Algorithm used by OLIGO for determining self-complementarity of oligonucleotides. A, the 5'-terminus of the oligonucleotide is progressively moved toward the 3'-terminus, and at each step, a determination of the number of base pairs is made. If a minimum length of continuous complementarity is set to 2, base pairing found at step 3 and 7 will be considered significant and will be stored in memory. No base-pairing occurs between A_6 and U_9 (step 7), since the minimum hairpin loop size is 3 (Freier *et al.*, 1986). B, calculation of ΔG for conformation 7 is shown as an example. ΔG_L, ΔG_R, ΔG_i, ΔG_{sym}, and ΔG_{loop} are free-energy increments, respectively, for left and right terminal mismatch, initiation, symmetry and loop; values are given in kcal/mol.

METHODS

Calculation of T_d. The following expression, employing nearest-neighbor thermodynamic values, was adopted from Freier *et al.* (5):

$$T_d = \frac{\Delta H}{\Delta S + R \times \ln(C/4)} - 273.15°C - t, \qquad \text{(ii)}$$

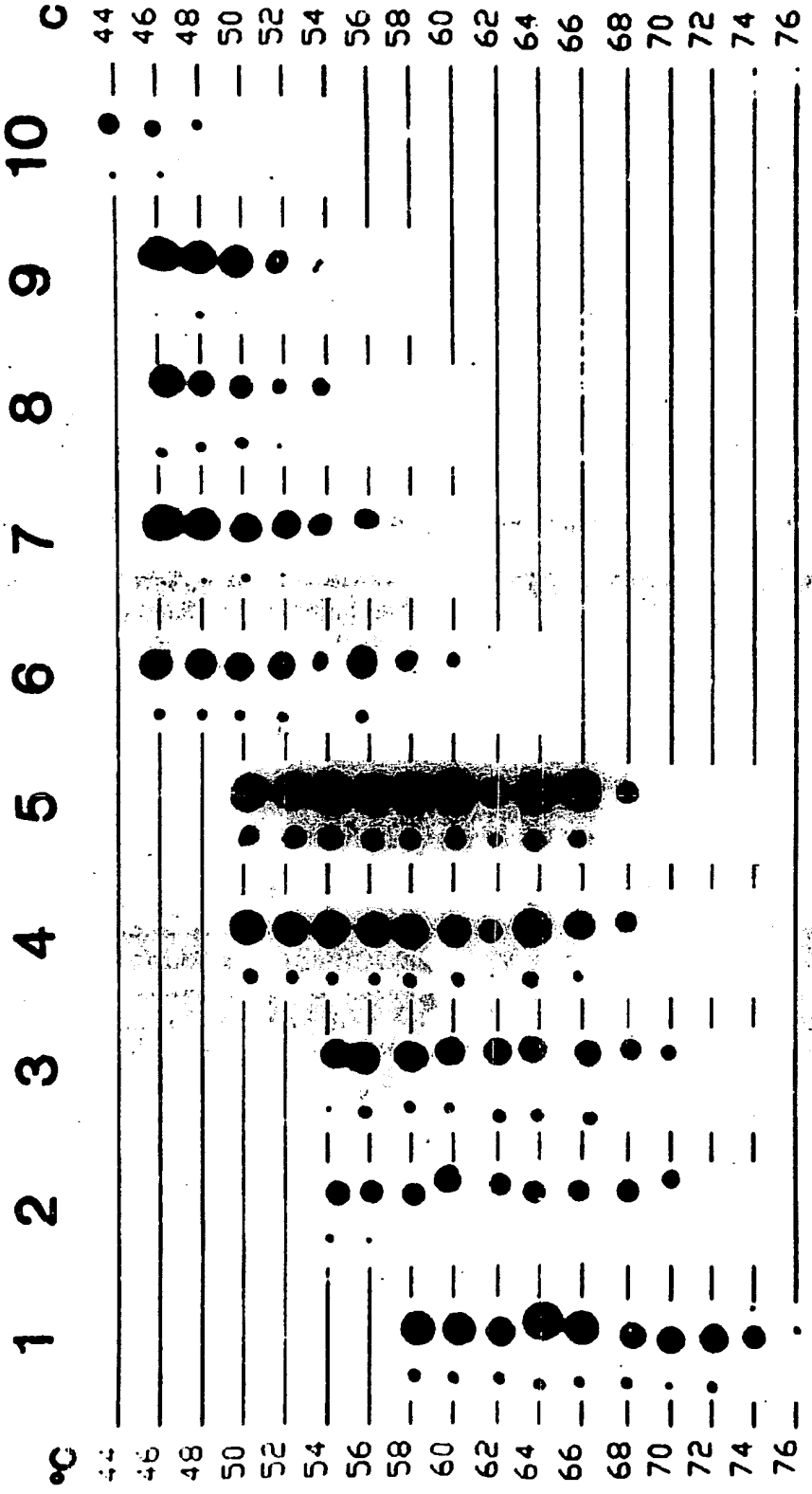
where ΔH and ΔS are the enthalpy and entropy for helix formation, respectively, R is molar gas constant [1.987 (cal/°C × mol)], and C is concentration of the probe. Thermodynamic parameters are those of Breslauer *et al.* (4) in the case of DNA and Freier *et al.* (5) for RNA. A novel constant, *t*, is introduced here and represents a temperature correction for filter hybridization, since the original work of Breslauer *et al.* (4) and Freier *et al.* (5) determined nucleic acid melting temperatures (T_m) in solution. In the experiments described here and by Suggs *et al.* (3), T_d was based on filter hybridization. A value of 7.6°C was empirically determined for *t* for the case where about 10 fmol of DNA is spotted on 1 mm^2 of filter area. The optimal hybridization temperature depends on the time for which filters are washed but is generally 5 to 10°C lower than T_d.

The program also calculates T_m, which is more useful for planning DNA amplification experiments. In such a case, the value *t* of the equation (ii) is equal to 0.

Determination of self-complementarity. The 5′-terminal sequence of a given oligonucleotide is aligned with the adjacent sequences as shown in Figure 1A. After a determination of base-pairing is made at the initial position (position 1), the 5′ terminus is repositioned one nt closer to the 3′ terminus and the oligonucleotides are checked for base-pairing again (position 2). If the length of continuous duplex is equal to or higher than a pre-set minimum, the position and ΔG, calculated as shown in Figure 1B, are stored in memory. The program stores in memory only up to 100 such positions, since it is not intended to predict secondary structure for large nucleic acids.

Determination of complementary regions. A given oligonucleotide is aligned with the 5′-terminus of the nucleic acid sequence and checked for base-pairing. If the length of continuous base-pairing is longer than a pre-set value, the position numbers of the fragments and length of double-stranded region are saved in memory. The oligonucleotide is then repositioned one nt closer to the 3′-terminus of the nucleic acid sequence. For PCR reactions, it is possible to check that primers are not complementary to each other with the same algorithm.

Determination of palindromes. The choice of an oligonucleotide which anneals to all or part of a palindromic sequence for the nucleic acid in question would cause problems for double-stranded sequencing. The program OLIGO uses the following algorithm to find palindromes. A variable *min* is defined as half the minimum size considered to be a palindrome. Beginning with position *min* of the nucleic acid sequence, a 'hairpin structure' is formed, and base-pairing between position *min* and *min* + 1 is checked. If complementarity is observed, positions *min* − 1 and *min* + 2 are similarly checked. This process is continued until a mismatch is found. If the length of continuous base-pairing exceeds *min*, the position and length are stored in memory. The process is then repeated beginning with position


°C
44
46
48
50
52
54
56
58
60
62
64
66
68
70
72
74
76
10
9
8
7
6
5
4
3
2
1

Tab.e I. Predicted and actual dissociation temperatures of selected oligonucleotides.

Number	Sequence	Length (nt)	Dissociation temperature, T_d (°C) experimental data[a]	predicted from: equation (i) and error[b]	equation (ii) and error
1	GTCGAACCGGAAACCACCCCT	21	72	68 (−0.7)	73.4 (1.4)
2	GTGCCCATCTGTTCTGTAGGGG	22	69	70 (4.3)	69.8 (0.8)
3	TCCGGTTCGACAGTCGCC	18	68	60 (−4.7)	69.1 (1.1)
4	CTGGATATGGTTGTACAGAGCCC	23	67	70 (6.3)	67.0 (0.0)
5	GGAGATCAGCCGCAGGTTT	19	67	60 (−3.7)	66.6 (−0.4)
6	CAGCGCCACATACATCAT	18	60	54 (−2.7)	59.3 (−0.7)
7	GACGCAGTCTCCTATGAG	18	55	56 (4.3)	53.7 (−1.3)
8	AAAGCAGTCCCATTCAT	17	53	48 (−1.7)	53.7 (0.7)
9	CAAAGGTGGAATAAACAT	18	52	48 (0.7)	51.5 (−0.5)
10	CCCAGTTTTAATATTTG	17	48	44 (0.7)	46.9 (−1.1)

[a] Experimental values of T_d were derived from densitometric scanning of the autoradiogram shown in Fig. 2.
[b] Error is a difference between empirical and theoretical melting temperature plus offset k.

$$\text{Error} = T_{exp.} - T_{theor.} + k,$$

where k is the sum of differences between the empirical and theoretical values divided by number of oligonucleotides tested:

$$k = \frac{\Sigma}{n}(T_{exp.} - T_{theor.})/n$$

and is equal to 3.3°C and 0°C for the equations (i) and (ii), respectively.

min + 1 of the nucleic acid sequence and continued until the entire sequence has been checked for palindromes.

Filter hybridizations and T_d determination. Either 2 or 20 fmol of plasmid or phage lambda DNA, in a volume of 0.5 μl, were spotted on MSI Nylon filters (Micron Separations Inc.) and hybridizations were carried out using 5'-^{32}P-labeled oligonucleotides with a specific activity of $10^8 - 10^9$ cpm/μg as previously described (6). Filters were washed twice at room temperature for 5 min and kept on ice-cold buffer until washed at a higher temperature, as specified in text, for 15 min. The filters were then subjected to autoradiography and the radioactivity quantitated by densitometry. The T_d was defined as the highest temperature at which significant amount of the radioactivity was retained on the filter. For the 2-fmol series, a significant amount was considered to be 10−33% of the strongest signal in the temperature sequence; for the 20-fmol series, this was considered to be 33−50% of the strongest signal. The T_d was calculated as the average of the two series.

Sequencing reactions. The single-stranded form of plasmid pTZ18R (Pharmacia) containing cDNA for protein synthesis initiation factor 4E was used (6). Sequencing reactions were performed using a modified T4 polymerase (Sequenase®, United States Biochemicals)

Figure 2. Hybridization of oligonucleotides to the target DNA. Two series of target DNA were spotted on the filter, 2 fmol (spots on the left of each numbered column) and 20 fmol (right). Numbers on the top refer to the oligonucleotides listed in the Table I. The temperature of the final wash is indicated. An autoradiogram is shown.

OLIGONUCLEOTIDE 6:

18-mer: 5' CAGCGCCACATACATCAT 3'
(above sequence is complementary to mRNA);
Hybridization temperature = 59.1°C [c=0.100 nM];
ΔG=-28.6 kcal/mol; ΔH=-132.3 kcal/mol; ΔS=-340.6 eu;

Duplex formation ("+" strand):

```
5' ATGATGTATGTGGCGCTG 3'
                ||||
          3' GTCGCGGTGTATGTAGTA 5'
```

Self-complementarity ("+" strand):

```
          5'  ATGA┐
              ||  T
  3' GTCGCGGTGTATG┘
```

Complementary fragments:

```
     1 CAGCGCCACATACATCAT 18 (457 to 440)
1.   ||||||||||||||||||
   457 GTCGCGGTGTATGTAGTA 440

     1 CAGCGCCACATACATCAT 18
2.        |||||||
   311 AGAAGGGTGTATCCGAGT 294
```

OLIGONUCLEOTIDE 10:

17-mer: 5' CCCAGTTTTAATATTTG 3'
(above sequence is complementary to mRNA);
Hybridization temperature = 46.7°C [c=0.100 nM];
ΔG=-24.1 kcal/mol; ΔH=-131.7 kcal/mol; ΔS=-353.5 eu;

Duplex formation ("+" strand):

```
       5' CAAATATTAAAACTGGG 3'
                 ||||||
3' GGGTCAAAATTATAAAC 5'
```

Self-complementarity ("+" strand):

```
       5'  CAAATAT┐
           ||  |  T
     3' GGGTCAAAA┘
```

Complementary fragments:

```
     1 CCCAGTTTTAATATTTG 17 (-311 to -327)
1.   ||||||||||||||||||
  -311 GGGTCAAAATTATAAAC -327

     1 CCCAGTTTTAATATTTG 17
2.   | | ||||||||||
  1590 GTGACAAAATTATACTT 1574

     1 CCCAGTTTTAATATTTG 17
3.      | ||||||||
   114 CCCTACAAATTATATCA 98

     1 CCCAGTTTTAATATTTG 17
4.     | | | | |||||||
  1874 GATTTATATATATAAAC 1858

     1 CCCAGTTTTAATATTTG 17
5.       |  |  |||||||
 -2838 CTCGCCTATGTATAAAC -2854

     1 CCCAGTTTTAATATTTG 17
6.       |    |||||||
  1212 TCTTAGTGATTATAATA 1196

     1 CCCAGTTTTAATATTTG 17
7.       |||||||     ||
  1602 ATGTCAAAACATGTGAC 1586
```

under the conditions supplied by the manufacturer. Primers were synthesized with an Applied Biosystems Model 380B DNA Synthesizer.

RESULTS

Duplex dissociation temperature. In order to compare the predictions of T_d calculated from Equations (i) and (ii), ten oligonucleotides ranging from 17 to 23 nt in length were used (Figure 2). Experimental T_d was determined from the retention of radioactivity at increasing temperatures as described in Methods. Differences between theoretical and experimental data are shown in the Table I. The mean errors for equations (i) and (ii) were 3.0°C and 0.8°C, and the largest errors were 6.3°C and 1.4°C, respectively. Equation (ii) was thus approximately four times more accurate than equation (i). Aside from the difference in predicted *vs*. experimental T_d values, it should be noted that experimental T_d values reported here are 3.3°C higher than the experimental values determined by Suggs *et al.* (3). This is likely to be due to a difference in methods of hybridization employed: Suggs *et al.* (3) washed filters for increasing lengths of time, as the temperature was raised, since the same filters were used succesively. In the present study, separate filters were used for each temperature, and all filters were washed for the same length of time, regardless of the temperature.

Probe specificity and secondary structure. An oligonucleotide probe should be not complementary to nucleic acid regions other than the target sequence. Checking the entire nucleic acid sequence for the possible formation of stable duplexes with a given oligonucleotide is important if the oligonucleotide is to be used as a sequencing primer, as discussed above. In order to illustrate the importance of the probe specificity, two sequencing primers, oligonucleotides 6 and 10 (Table I) were compared for their ability to prime sequencing reactions (Figure 3). Even though both oligonucleotides formed perfect duplexes with the test plasmid and differed by only one nt in length, oligonucleotide 6 was strikingly more specific as primer. Comparison of these oligonucleotides using OLIGO revealed that oligonucleotide 6 formed a single non-target site duplex (7 nt), whereas oligonucleotide formed six such duplexes of up to 10 nt in length. Perhaps also significant is the fact that two of the latter duplexes involved the 3'-terminus of the oligonucleotide. At the relatively low temperature of the sequencing reaction (20°C), oligonucleotide 10 apparently forms duplexes with these non-target sequences resulting in its poor specificity as a sequencing primer. Neither oligonucleotide was self-complementary, indicating that secondary structure was not responsible for the poor results with oligonucleotide 10. Similarly, potential problems due to dimer formation were excluded.

DISCUSSION

Several methods of calculating hybridization temperatures for oligonucleotide probes are known. A commonly used one is that of Suggs *et al.* (3), where T_d calculation is based on the number of AT and GC base pairs. A more precise method for determination of duplex-melting temperature is based on nearest-neighbor thermodynamic parameters (4, 5). The major drawbacks of using these parameters are first, that they apply to solution

Figure 3. Comparison of the two sequencing primers, oligonucleotides 6 and 10. An autoradiogram of the sequencing gel is shown on the left. On the right is a full list of the target DNA fragments complementary to the primers (with continuous complementarity of 7 or more nt). The possible dimer formation, self-complementarity, and the thermodynamic parameters calculated by OLIGO are shown.

hybridization, and second, that calculation of melting temperature is too laborious to be done by hand for long sequences. The results presented here demonstrate that with a suitable computer program, the nearest-neighbor approach can be adapted for filter hybridization methods. This technique is significantly more accurate than the method based on numbers of AT and GC pairs (3). We found that screening phage libraries at a density of 500 plaques per cm^2 was specific and reproducible when the temperature of the final wash and/or temperature of hybridization was 5°C lower than T_d calculated according to the equation (ii). At the time that an oligonucleotide is being selected on the basis of the maximal T_d, the program OLIGO can be used to determine its suitability based on self-complementarity, specificity and absence of a palindrome in the target site.

The algorithm employed in this program finds all possible duplexes, independent of whether they overlap with each other, unlike several other algorithms used in determination of RNA secondary structure (7–9). Further analysis of the duplexes is needed to distinguish simple hairpin loops from bulge loops, unbalanced interior loops, pseudoknots or other structures.

Oligonucleotides complementary to palindromic sequences are poor sequencing primers not only because they can form dimers, and therefore not anneal to the target sequence efficiently, but also, when dsDNA is sequenced, the synthesis would occur simultaneously in both directions. In addition, the presence of palindromic sequences is frequently of interest for reasons other than sequencing. Thus it is important to check the whole sequence file for palindromes. This is possible with OLIGO since the algorithm is very fast. The total number of base comparisons is variable and depends on the sequence, but is always close to $n+0.3\times n$, where n is the number of base pairs in the DNA sequence.

OLIGO is designed to find optimal sequencing primers, PCR primers and hybridization probes. Each of the parameters determined (duplex stability, specificity, self-complementarity of oligonucleotides, presence of palindromes, ability of the probes to form dimers) are checked independently. Selection of the probe is made by the investigator. The program is written in Turbo C (Borland International) for IBM PC-compatible computers and is available upon request.

ACKNOWLEDGEMENT

This work was supported in part by Grant No. GM 20818 from the National Institute of General Medical Sciences.

*To whom correspondence should be addressed

REFERENCES

1. Berent, S.L., Mahmoudi, M., Torczynski, R.M., Bragg, P.W., and Bollon, A.P. (1985) BioTechniques 3, 208–220.
2. Ausubel, F.M. (1987) In Ausubel, F.M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (eds.), Current Protocols in Molecular Biology. Greene Publishing Associates and Wiley-Interscience, pp. 6.4.1–6.4.10.
3. Suggs, S.V., Hirose, T., Miyake, E.H., Kawashima, M.J., Johnson, K.I., and Wallace, R.B. (1981) In D.D. Brown (ed.), ICN-UCLA Symp. Dev. Biol. Using Purified Genes. Academic Press Inc., New York, Vol. 23, pp., 683–693.
4. Breslauer, K.J., Frank, R., Blocker, H., and Marky, L.A. (1986) Proc. Natl. Acad. Sci. USA 83, 3746–3750.
5. Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T., and Turner, D.H. (1986) Proc. Natl. Acad. Sci. USA 83, 9373–9377.
6. Rychlik, W., Domier, L.L., Gardner, P.R., Hellmann, G.M., and Rhoads, R.E. (1987) Proc. Natl Acad. Sci. USA 84, 945–949.

7. Nussinov, R., and Jacobson, A.B. (1980) Proc. Natl. Acad. Sci. USA 77, 6309–6313.
8. Zucker, M., and Stiegler, P. (1981) Nucleic Acid Research 9, 133–148.
9. Williams, A.L., and Tinoco, I. (1986) Nucleic Acid Research 14, 299–315.